*Bureau canadien
des brevets*

Certification

*Canadian Patent
Office*

Certification

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,331,116,** on January 15, 2001, by **CHENOMX, INC**, assignee of David S. Wishart, Russ Greiner, Jonathan Schaeffer, Brent Lefebvre, Ian Parsons, Tim Rosborough and Warren Wong, for "Compound Identification and Quantitation in Liquid Mixtures - Method and Process Using an Automated Nuclear Magnetic Resonance Measurement System".

Agent certificateur/Certifying Officer

November 13, 2003

Date

COMPOUND IDENTIFICATION AND QUANTITATION IN LIQUID MIXTURES -- METHOD AND PROCESS USING AN AUTOMATED NUCLEAR MAGNETIC RESONANCE MEASUREMENT SYSTEM

I.   ABSTRACT

This invention is a process and method for high-throughput automated compound identification and quantitation ("CIQ") of multiple compounds present in biological fluids that include, but are not limited to, human blood and urine using nuclear magnetic resonance spectroscopy ("NMR").

COMPOUND IDENTIFICATION AND QUANTITATION IN LIQUID MIXTURES --
METHOD AND PROCESS USING AN AUTOMATED NUCLEAR MAGNETIC
RESONANCE MEASUREMENT SYSTEM

The general steps involved in the CIQ system include the following:

a) a biological fluid sample is initially "doped" with a small quantity of a chemically inert pH indicator and an appropriate chemical shift reference standard,

b) the biological fluid sample is then loaded into a commercially available auto-sampler that takes a portion of the sample for analysis;

c) the sample is forced through a small diameter tube by a liquid syringe into a commercially available spectrometer;

d) electromagnetic energy in the radio frequency range from the spectrometer is applied to the biological fluid sample using Nuclear Magnetic Resonance ("NMR") to cause a temporary change in the magnetic properties of the protons contained in the sample;

e) as the protons revert to their original or ground state, they absorb characteristic radio frequency energy that is detected and recorded by the spectrometer as data that represents the Free Induction Decay ("FID") of the measured emissions;

f) the FID data is converted into a spectral profile using sophisticated hardware and proprietary software that perform Fourier transforms on the data;

g) the data is then processed using a software program that converts the transformed data into a trace file of points in a X-Y plane that represents the spectral profile of the data recorded by the spectrometer;

h) a "peak picking" software program is then used to identify the location of the peak of the internal standard in the spectral profile;

i) a software algorithm called "Nfit" is then used on the spectral profile to refine the location of the internal standard peak, references that point as 0 ppm and assigns Lorentzian parameters to the internal standard peak;

j) the remaining peaks in the spectral profile are then initially identified and quantified in relation to the internal standard peak using a peak picking program;

k) the invention has a database of NMR spectral profiles of known standard compounds at different pH values;

l) the pH value of the biological fluid sample is determined using the pH reference compound, in which the position of its peaks are affected by pH in a known way.

m) using the database of known standard compounds that have precisely the same pH value of the biological fluid sample, the peak locations of the known standard compounds are compared and matched with the detected peak locations of the sample;

n) a "non-negative linear least squares" algorithm is then used to calculate the best linear combination of the database compounds that optimally matches the spectral profile of the biological fluid sample, this allows all the compounds in the biological sample to be simultaneously identified and quantified;

o) the invention, by virtue of its nearly complete level of automation, can perform a quantitative spectral analysis of a liquid mixture containing hundreds of components in a matter of minutes as opposed to hours or days using traditional manual methods.

II. COMPOUND IDENTIFICATION AND QUANTITATION IN LIQUID MIXTURES - INTRODUCTION

There are other methods available to identify and quantify biological compounds in liquid mixtures. Almost all of these approaches require some kind of initial compound separation step (gas chromatography, electrophoresis, liquid chromatography) followed by some kind of spectral detection or identification step (mass spectrometry, ultraviolet spectroscopy or infrared analysis). These methods are expensive, manually intensive and require considerable technical expertise and time. More recently, NMR spectroscopy has been proposed as an alternative approach to liquid mixture analysis because it does not require chromatographic compound separation. In NMR spectroscopy, radio frequency (RF) electromagnetic radiation is applied to a mixture of organic compounds. This allows one to extract and measure the characteristic RF absorption frequencies of the protons belonging to the chemical compounds found in the mixture. The measured data is then Fourier transformed to obtain a spectral profile in which different compounds display well-separated sets of peaks or RF absorption bands. The spectral profile is then **manually** analyzed by an NMR expert using tables of known or partially known chemical shifts of previously measured pure compounds. In this way it is possible to identify some of the chemical components of the mixture being analyzed. While this NMR approach is quite appealing in that it avoids chromatographic separation, it is nevertheless quite slow, relatively inaccurate and somewhat limited by the expertise of the NMR analyst

The invention described herein sets forth a method and process that is novel in the automated, user-independent techniques that are used to accurately identify and quantitate large numbers of components in the complex mixtures commonly found in biological fluids. The invention described herein also sets forth an improvement over traditional methods of NMR analysis by automating the steps required to acquire, process and compare NMR spectra and by dramatically increasing the speed and throughput capacity of the process such that an accurate spectral analysis of complex liquid mixtures can be performed in minutes as opposed to hours.

III. LIQUID MIXTURE IDENTIFICATION AND QUANTITATION – SUMMARY

Functional Flowchart:

The entire method and process is represented by the attached flowchart diagram that describes the following:

1. Read and store FID from spectrometer.

2. Preprocessing spectrum.
   - (I)   Spectrum Handling
   - (II)   Baseline Correction
   - (III) Find Internal Standard
   - (IV)   pH determination
   - (V)   Peak picking

3. Read and Preprocess Compound Database
   - (I)   Read database and incorporate pH value into pH curve equations
   - (II) Create new list of peak positions for all database compounds

4. Quantitation and Identification
   - (I)   Perform an initial assignment and quantitation of peaks
   - (II)   Apply "Wiggle" and linewidth adjustment to Compound Database peaks
   - (III) Apply Generalized Linear Least Squares algorithm to refine assignment

5. Report results into the table of "Compounds Detected and Associated Concentrations".

IV.   DESCRIPTION OF THE CIQ PROCESS

The function of each element of the process is described as follows:

1. Read spectrum:
   The FID data of a biological fluid sample is recorded using a commercially available NMR spectrometer and then converted by applying a Fourier transformation of the recorded FID data.

2. Preprocessing Spectrum

a) Spectrum Handling. The Fourier transformed data is processed by software to convert the data into a trace file. Line broadening is applied to improve the quality of the data. A "drift" correction is performed on the data by adjusting the two extremes of the baseline to "zero". The data is processed again to suppress the presence of water in the spectrum. Additional processing is performed on the data to make the peaks in the data as symmetrical as possible (this is called phasing). Then, a collection of data points are extracted from the spectrum to form a spectral trace file and saved separately. Finally, all of the FID data files are archived in another file for future reference. The entire spectrum handling process is performed automatically using a combination of pre-existing and newly developed software.

b) Baseline Correction. The data is automatically baseline corrected by modelling the noise in the data to yield a straight horizontal baseline without affecting the peak positions.

c) Find Internal Standard. The data is then automatically processed to identify the internal standard, DSS (3-[trimethylsilyl]-1-propanesulfonic acid). A "Nfit" algorithm is applied to the data to find the position of the DSS peak centre and then the data is processed to reference the new centre of the DSS peak to 0 ppm.

d)   pH Determination.   The data is then automatically processed to identify the peak corresponding to the pH reference compound (TSP – trimethylsilysl-1-propanoic acid). The location of the peak in relation to DSS (0 ppm) is identified and used in a formula to calculate the actual pH of the sample being analyzed.

e)   Peak Picking.   The spectrum from the liquid mixture is then peak picked using a combination of two specially developed peak pickers. This creates an initial set of peak positions and intensities which is used to help guide the subsequent compound identification and quantitation steps.

3.   Read and Preprocess Compound Database.

The Compound Database is a proprietary database containing the peak positions, peak width, peak assignments and peak intensities of nearly 200 small molecule compounds known to exist in biological fluids. The database also includes peak position tolerances ("wiggle factors") and previously measured pH titration curves for pH sensitive peaks. In the preprocessing step the pH value of the liquid mixture (see step 2d) is incorporated into the pH titration equations for each pH sensitive compound in the database. A new peak list is then automatically generated for all the compounds in the database at that pH. This pH-adjusted list serves as the database for the subsequent identification and quantitation step.

4.   Identification and Quantitation.

Using the list of peaks previously identified in Step 2e, a rapid peak comparison algorithm ("Identify") is run to identify and approximately quantify an initial subset of compounds in the liquid mixture sample. The algorithm compares and matches the peaks found in the sample spectrum with the pH-corrected peaks listed in the Compound Database. Using a simple scoring system and threshold analysis the algorithm is able to identify a significant portion of the compounds in the mixture in a matter of seconds. After this initial pass, the exact identity and quantity of the compounds in the mixture is refined using a second, more detailed processing step. In this refinement step the spectral linewidths in the sample spectrum are first calibrated relative to the DSS peak. Then two separate tests are applied to the sample spectrum to further ready it for spectral fitting and refinement. The first test is a "wiggle" algorithm that adjusts the database compound peaks in one direction or another until the peak matches those in the sample spectrum. A second test is a filtering algorithm that identifies and removes isolated peaks along with other peaks elsewhere in the spectrum associated with the same compounds. The results of both tests are passed to a modified non-negative linear least squares algorithm which precisely fits the observed sample spectrum with the derived spectra of its pure components. In this way the exact composition and concentration of all the compounds in the mixture is simultaneously determined.

5.   Results Reported

The results of the identification and quantitation process are reported in the table of the "Compounds Detected and Associated Concentrations".

V.   CIQ HARDWARE OVERVIEW

The CIQ system is comprised of the following hardware components:

a)   Varian NMR Spectrometer equipped with a shielded Oxford 400 MHz magnet;

b)   Varian Mercury console;

c)   Sun Computer Workstation; and

d)   Modified Gilson/Varian Versatile Automated Sample Transport (VAST) Auto-sampler.

e)   Associated plumbing and valves for liquid handling

VI.   CIQ SOFTWARE OVERVIEW

The CIQ system is comprised of the following proprietary software modules:

a)   Read Spectrum of FID data;

b)   Preprocessing of Spectrum – spectrum handling, baseline correction, find internal standard, pH determination and peak picking;

c)   Read Compound Database;

d)   Preprocess Compound Database;

e)   Identification and Quantitation; and

f)   Archiving of Processed Data.

{EM101069.DOC;1}

Overview of CIQ (Compound Identification and Quantitation)

1. Read spectrum
- the FID data is recorded and converted using Fourier Transform

2. Preprocessing spectrum
(I) Spectrum handling
(II) "baseline correction"
(III) DSS reference
(IV) pH reference

(I) Spectrum Handling
- autoprocessing software that converts FT data in a trace file

A. Set linewidth to 0.5 and re-FT the data
- to eliminate glitches and increase the quality of the data.

B. perform "drift correction
- sets the two extremes of the baseline to 0.

C. "nudge water"
- optimize water suppression
- minimize the area of the peak within a 0.5 ppm wide region centered on the water signal
- ideally reduces the area to zero

Find large water signal in the region 4.49 to 4.99

Using an adaptation of Brent's Method, SSLSFRQ value is adjusted
- Choosing successive satsfrq values, performing FT and comparing to an idealized parabolic curve until convergence occurs

F. "aph0"
- phases the spectrum and makes the peaks as symmetrical as possible

E. Collection of data points are extracted and saved to a set of external files

D. Package all the FID's and export them into a "tar" format file for archiving

(II) Baseline Correction
- "models" the noise to yield a straight horizontal baseline without affecting peak positions

A. take an average of a sliding window across the spectrum, resulting in a "baseline model"
- sliding window consists of 51 data points

B. average the spectral "noise" (deviations in baseline due to electronic interference)
- at an extreme end of the spectrum, where no true signals exist

C. Identify "noise" versus real peaks
- peaks are identified as anything with an intensity greater than 10X the standard deviation of noise
- this is performed as the sliding window moves across the spectrum

D. identified peaks are subtracted from the baseline, resulting in "holes" (peak width+51 data points)
- the software compensates by drawing a straight line across the holes
- this results in a complete "baseline model"

E. baseline "model" is subtracted from the original spectrum (containing peaks)
- results in a straight horizontal baseline with the chemical shifts in their original positions that occurred during sample acquisition

3. Read Compound Database
- Read database compound entries: peak positions (in ppm), "Wiggle" factors for each peak

4. Preprocess Compound Database

I) pH value of the sample is incorporated into the pH curve equations of each database compound

II) New list of peak positions for all database compounds is created
- Contains data on the position of the centers of each peak in the mixture sample, relative linewidths of each peak to the linewidth of DSS

5. Quantitation
- Using relative linewidths of each peak to DSS

I) Assign peaks in sample spectrum to compounds in the database

(II) Apply "Wiggle" to database compound peaks
- Moves the database compound peaks, 1 pixel in either direction, gradually increasing the pixel count until the peak is matched to those in the sample spectrum.

(III) Peak Picker
- Identifies isolated peaks and removes those peaks, as well as other peaks elsewhere in the spectrum, associated with the same compounds

6. Results reported in the table of "Compounds Detected and Associated Concentrations"

(III) Find Internal Standard
DSS (3-(trimethylsilyl)-1-propanesulfonic acid)

(IV) pH determination
- identifies the downfieldmost imidazole peak (pH reference)

A. Identify the upfieldmost peak as DSS
- Determines the average noise level across the first 600 data points
- Assumes anything over a certain multiplier of the average noise value is a peak

A. Identify the imidazole peak
- Using the amplitude of DSS, the software predicts the theoretical amplitude of imidazole
- Assumes that a peak located within a certain range (7.66 to 8.66 ppm) with the same amplitude expected of imidazole (with allowable error) is imidazole

B. Applies "nfit" algorithm
- Finds position of the DSS peak center, peak amplitude and width at half the maximum amplitude

1. gradient descent
- Find optimal or "true" values of A,c,w

Starting with initial guesses for amplitude (A), peak center (c), and half width (w) values, a formula for a Lorentzian function f and error formula E for signal s from f are used to iteratively determine better values for A,c,w that minimize E.

Minimum error value = "true" A,c,w values

B. Applies "nfit" algorithm
- Finds position of the peak center, peak amplitude and width at half the maximum amplitude

2. conjugate gradient
- Optimization to gradient descent method
- Uses orthogonal derivatives to calculate higher magnitude changes to A,c,w at each iteration
- Reduces the amount of iterations necessary to arrive at a minimum error value

C. References the new center of the pH referencing imidazole peak relative to DSS

C. References the new center of the DSS peak to 0 ppm

D. pH determination
- the location of imidazole (in ppm) is incorporated into the pH formula derived by Dr. Brian Sykes

(III) Generalized Linear Least Squares (NLLS)
- software determines which compounds from the database are present in the mixture during step SD(I & II).
- software uses LLS to determine the best linear combination of peak amplitudes (concentrations) of those identified compounds to match the sample to the spectrum
- iteratively adjusts the amounts of identified compounds in the "fitted" spectrum until the difference (error) reaches a minimum
  - this is performed by setting the error equation to 0
  - Non-negative Linear Least Squares (NNLS) is applied to each of the clusters of each database compound. Theoretical concentrations of each compound cluster are determined using a complex matrix.
  - Algorithm sets the minimum concentration of each compound as 0.
  - Algorithm sets the maximum concentration of each compound as the minimum value of the calculated clusters.

A) Triangle Method
- Uses a "moving triangle" strategy
- No filtering technique

B) Galactic Method
- Uses a "moving triangle" strategy
- Performs a set of sophisticated filtering and refinement techniques on the resulting set of peaks, including: smoothing, edge detection, minimum edge slope filtering, maximum peak group separation filtering, and maximum shoulder peak ratio filtering, and finally peak rejection parameters.

C) Results from both methods are merged, and redundant peaks are eliminated